

# Prädiktion mit Alters-Perioden-Kohorten- Modellen

BACHELORTHESIS



LUDWIG-MAXIMILIANS-UNIVERSITÄT

MÜNCHEN

JONAS SCHWEISTHAL

betreut von  
Prof. Dr. VOLKER SCHMID

August 2019

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>2</b>
<b>2</b>	<b>Theorie hinter bayesianischen APC-Modellen</b>	<b>4</b>
2.1	Problematik in der Identifizierbarkeit . . . . .	4
2.2	Random Walk Priorverteilungen . . . . .	5
2.3	Zusätzliche Heterogenität . . . . .	6
2.4	Prädiktion . . . . .	7
2.5	Methoden zur Parameterschätzung . . . . .	8
<b>3</b>	<b>Modellierung von Todesfällen durch Krebs im Verdauungstrakt</b>	<b>9</b>
3.1	Beschreibung der Daten . . . . .	9
3.2	Anwendung von APC-Modellen . . . . .	10
3.3	Modellevaluierung . . . . .	14
3.4	Einbinden von Kovariablen in APC-Modelle . . . . .	15
3.5	Weitere Möglichkeiten zur Modellierung . . . . .	17
<b>4</b>	<b>Retrospektive Prädiktion</b>	<b>19</b>
4.1	Methodik . . . . .	19
4.2	Bewertung der Vorhersageperformance . . . . .	20
<b>5</b>	<b>Vorhersage für zukünftige Jahre</b>	<b>23</b>
5.1	Modellwahl für die Prädiktion . . . . .	23
5.2	Interpretation der Prädiktionen . . . . .	24
<b>6</b>	<b>Konklusion</b>	<b>29</b>

# Kapitel 1

## Einleitung

Die Entwicklung einer Bevölkerung in Bezug auf verschiedene Faktoren zu ermitteln und darzustellen, ist in vielen unterschiedlichen Bereichen ein großes Anwendungsgebiet der Statistik. Zu diesen Faktoren zählt oft das Eintreten eines gewissen Ereignisses, beziehungsweise der Zeitpunkt des Eintretens. Für demographische Untersuchungen ist beispielsweise der zeitliche Verlauf der Todesfälle in einem Staat ein wichtiges Merkmal, um die jeweilige Bevölkerungsanzahl und deren Altersstruktur über die Jahre hinweg zu analysieren. Unter anderem in Hinblick auf die staatliche Rentenversicherung und dem damit einhergehenden Generationenvertrag spielen diese Kennzahlen eine wichtige Rolle (Ebert, 2018). Auch im Gesundheitswesen und speziell in der Epidemiologie werden Sterberaten, ausgelöst durch bestimmte Krankheiten, und ihre Entwicklung genauer betrachtet, um deren aktuelle beziehungsweise historische Bedeutung festzustellen und gegebenenfalls mögliche Ursachen zu finden.

In vielen Fällen sind der Bevölkerungsbestand und die Sterbefälle nach Altersgruppen in einem zeitlichen Verlauf gegeben, aber interessante Kovariablen stehen nicht zur Verfügung. Das kann einerseits an fehlenden Erhebungen liegen, andererseits ist oft nicht bekannt, welche Variablen als sinnvoller Einfluss in ein Modell aufgenommen werden könnten. In solchen Fällen werden häufig Alters-Perioden-Kohorten Modelle (im Folgenden: APC-Modelle) verwendet. Hierbei wird die Sterberate in Abhängigkeit ihres zeitlichen Verlaufs untersucht, welcher in drei verschiedene Effekte aufgeteilt wird: Das jeweilige Alter zum Zeitpunkt des Todes, die Periode, also das jeweilige Todesjahr, und die Kohorte als Geburtsjahr der Verstorbenen. Durch die Unterteilung in diese Effekte können ohne weitere Daten auch Einflüsse durch nicht erhobene oder nicht verwendete Kovariablen modelliert beziehungsweise extrahiert werden. Ein Beispiel hierfür wäre die Sterblichkeitsrate durch Kinderkrankheiten, bei denen das Alter eine wichtige Rolle spielt. Der Periodeneffekt hingegen könnte sich bei Krebstoten aller Altersklassen in Gebieten nach einer atomaren Verseuchung zeigen, während der Kohorteneffekt sich unter anderem durch Mangelernährung von Kindern nach Kriegen auf verschiedene Todesarten auswirken kann (Schmid und Held, 2007).

Diese Arbeit beschäftigt sich näher mit den von Berzuini und Clayton (1994) und Besag u. a. (1995) beschriebenen bayesianischen APC-Modellen. Zur Verbesserung

der Schätzung und Prädiktion werden dabei die Alters-, Perioden- und Kohorteneffekte geglättet. Zudem müssen für die Vorhersagen der Effekte keine starken parametrischen Annahmen getroffen werden und unstrukturierte Heterogenität kann in das Modell mit aufgenommen werden. (Knorr-Held und Rainer, 2001)

Im Folgenden werden als laufendes Beispiel für die praktische Anwendung und die Prädiktion mit APC-Modellen Daten zur Sterblichkeit durch Karzinome im Verdauungstrakt in Deutschland von 1998 bis 2017 verwendet (für genauere Informationen zu den einzelnen Krebsarten siehe Deutsches Institut für Medizinische Dokumentation und Information DIMDI (2018)). Die Daten werden vom statistischen Bundesamt in der Gesundheitsberichterstattung der Länder bereitgestellt (siehe Statistisches Bundesamt DeStatis (2019)). Für jedes der 20 Jahre stehen die Bevölkerung und die Todesfälle durch Karzinome im Verdauungstrakt, unterteilt nach Geschlecht und Altersgruppen mit einer Breite von jeweils 5 Jahren zwischen einem Alter von 25 bis 89 Jahren, zur Verfügung. Sprich: Es ergeben sich 13 verschiedene Altersgruppen mit den Stufen 25-29, 30-34, ..., 85-89 Jahre. Von Schmid und Held (2004) wurde unter anderem schon die Sterblichkeit durch Magenkrebs, welche einen Großteil der Sterblichkeit durch Krebs in den Verdauungsorganen ausmacht, für Daten aus Westdeutschland von 1976 bis 1990 mit APC-Modellen analysiert und mögliche Einflussfaktoren angesprochen. Daran orientiert wird in dieser Arbeit untersucht, wie sich weitere Variablen in die Modellierung einbinden lassen und inwieweit sie eine Verbesserung, vor allem bezüglich der Prädiktion, darstellen können.

Alle Auswertungen werden mit der Statistik-Software *R* (R Core Team, 2018) erstellt. Für die Schätzungen der APC-Modelle, Vorhersagen und Generierung der Kennzahlen zur Modellevaluierung wurde das *R*-Package *bamp* verwendet (siehe auch Schmid und Held (2007)).

Diese Arbeit ist folgendermaßen untergliedert: Kapitel 2 behandelt die theoretischen Grundlagen der bayesianischen APC-Modelle und deren mögliche Erweiterungen durch zusätzliche Heterogenitätsparameter. In Kapitel 3 werden die verschiedenen Modellvariationen mittels *bamp* auf den Datensatz der Tode durch Karzinome im Verdauungstrakt angewendet, verglichen und evaluiert. Hierbei wird auch auf die Möglichkeit eingegangen, Perioden- oder Kohorteneffekte durch Einflüsse von erklärenden Variablen zu ersetzen. In Kapitel 4 wird rückblickend die tatsächliche Vorhersagekraft einzelner Modelle in der Vergangenheit betrachtet, indem die Beobachtungen der letzten Jahre nicht mit in die Modellierung aufgenommen werden und als Kontrolldaten fungieren. In Kapitel 5 werden anhand von ausgewählten Modellen Prognosen der Sterberaten für die Jahre 2018 bis 2022 präsentiert. Kapitel 6 fasst abschließend die bisherigen Erkenntnisse zusammen und geht dabei auf Probleme bei dem Ersetzen von Effekten durch Kovariablen ein.

# Kapitel 2

## Theorie hinter bayesianischen APC-Modellen

Der Inhalt des folgenden Kapitels orientiert sich, wenn nicht anders ausgewiesen, an den Artikeln von Knorr-Held und Rainer (2001) und Schmid und Held (2007).

Im Folgenden steht  $\eta_{ij}$  für die Chance des Sterbens für die Altersgruppe  $i$ , mit  $i = 1, \dots, I$  im Jahr  $j$ , mit  $j = 1, \dots, J$ . In unserem Beispiel gibt es  $I = 13$  Altersgruppen und  $J = 20$  Jahre. Der Index für die Kohorte  $k$ , mit  $i = 1, \dots, K$  muss für die jeweilige Chance nicht extra betrachtet werden, da er sich durch die Abhängigkeit aus den anderen Indizes als

$$k = k(i, j) = (I - i) + j \quad (2.1)$$

ergibt, vorausgesetzt Alter und Periode werden auf der gleichen Zeitskala gemessen, wie beispielsweise bei Ein-Jahres-Altersgruppen und jährlicher Erhebungen. Da für den Datensatz des Krebs im Verdauungstrakt allerdings Altersgruppen mit der Breite von jeweils 5 Jahre bei jährlicher Erhebung vorliegen, muss die Definition des Kohortenindex wie folgt angepasst werden:

$$k = k(i, j) = 5 \cdot (I - i) + j \quad (2.2)$$

Für die Modellierung der Todesfälle durch den Krebs  $y_{ij}$  wird eine Binomialverteilung  $y_{ij} \sim B(n_{ij}, \pi_{ij})$  mit den Parametern  $n_{ij}$  (Populationsgröße) und  $\pi_{ij}$  (unbekannte Sterbewahrscheinlichkeit) angenommen. Die Sterbechance  $\eta_{ij}$  als Logit der Sterbewahrscheinlichkeit  $\pi_{ij}$  wird über den Intercept  $\mu$ , den Alterseffekt  $\theta_i$ , den Periodeneffekt  $\phi_j$  und den Kohorteneffekt  $\psi_k$  dargestellt,

$$\eta_{ij} = \log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \mu + \theta_i + \phi_j + \psi_k. \quad (2.3)$$

### 2.1 Problematik in der Identifizierbarkeit

Um Identifizierbarkeit zu gewährleisten, müssen die Parameter folgende Beschränkung erfüllen:

$$\sum_i \theta_i = \sum_j \phi_j = \sum_k \psi_k = 0 \quad (2.4)$$

Aufgrund der linearen Abhängigkeit in (2.3) sind die Parameter jedoch trotzdem nicht eindeutig identifizierbar. Hierbei muss beachtet werden, dass der Kohortenparameter bei gleicher Zeitskala für Alter und Periode nicht einzelne Geburtsjahrgänge darstellt, sondern jeweils zwei Jahre. Somit überlappen sich die aufeinander folgenden Kohortenparameter jeweils um ein Jahr, wie unter anderem Holford (1983) festgestellt hat. Die Chance  $\eta_{ij}$  (und somit auch die Wahrscheinlichkeit  $\pi_{ij}$ ) bleibt für jede lineare Transformation mit beliebigem  $c$

$$\begin{aligned}\theta_i &\rightarrow \theta_i + c \left( i - \frac{I+1}{2} \right), \\ \phi_j &\rightarrow \phi_j - c \left( j - \frac{J+1}{2} \right), \\ \psi_k &\rightarrow \psi_k + c \left( k - \frac{K+1}{2} \right)\end{aligned}\tag{2.5}$$

unverändert für alle möglichen Werte für  $i$  und  $j$ . Das trifft auch für das Modell unserer Daten mit unterschiedlichen Zeitskalen für Alter und Periode zu, in dem beispielsweise zwei aufeinander folgende Kohorten  $\psi_r$  mit den Geburtsjahren 1915 bis 1920 und  $\psi_{r+1}$  mit den Geburtsjahren 1916 bis 1921 wären. Die lineare Transformation für den Alterseffekt ergibt sich in diesem Fall als

$$\theta_i \rightarrow \theta_i - 5 \cdot c \left( i - \frac{I+1}{2} \right).\tag{2.6}$$

Folglich sind die Chance und die Wahrscheinlichkeit immer eindeutig identifizierbar und interpretierbar, was aber nicht für die Alters-, Perioden- und Kohorteneffekte gilt. Hier können nur nicht-lineare Trends, wie zum Beispiel Wendepunkte, in den einzelnen Effekten getrennt interpretiert werden (siehe dazu auch Holford (1983) und Berzuini u. a. (1993)).

## 2.2 Random Walk Priorverteilungen

Bei bayesianischen Versionen der APC-Modelle werden für die Parameter von Alter  $\theta_i$ , Periode  $\phi_j$  und Kohorte  $\psi_k$  üblicherweise in Anlehnung an Berzuini und Clayton (1994) Random Walk (RW) Priorverteilungen der ersten oder zweiten Ordnung verwendet. Für den Intercept  $\mu$  wird ein flacher Prior angenommen. Für den Prior gibt es zwei verschiedene Möglichkeiten: Die erste ist der Random Walk erster Ordnung (RW 1), der von einem über die Zeit hinweg konstanten Trend ausgeht. Als zweite kann der Random Walk zweiter Ordnung (RW 2) angewendet werden, der von einem linearen Trend ausgeht.

Die unterschiedlichen Priorverteilungen bringen gleichzeitig unterschiedliche stochastische Restriktionen mit sich. Beim RW 1 werden die ersten Differenzen der Parameter für Alter, Periode und Kohorte als unabhängige, normalverteilte Zufallsvariablen angenommen. Ein dementsprechender geglätteter Prior für das Alter ergibt sich als

$$p(\theta|k) \propto \exp\left(-\frac{\kappa}{2} \sum_{i=2}^I (\theta_i - \theta_{i-1})^2\right), \quad (2.7)$$

für die

$$\theta_i \sim N(\theta_{i-1}, \kappa^{-1}), \quad \text{für } i = 2, \dots, I \quad (2.8)$$

gilt, wobei für  $\theta_1$  ein flacher Prior angenommen wird und  $\kappa$  als Präzisionsparameter fungiert.

Im Vergleich dazu werden beim RW 2 die zweiten Differenzen der Parameter als unabhängige, normalverteilte Zufallsvariablen angenommen. Somit stellt sich ein geglätteter Prior als

$$p(\theta|k) \propto \exp\left(-\frac{\kappa}{2} \sum_{i=3}^I (\theta_i - 2\theta_{i-1} + \theta_{i-2})^2\right), \quad (2.9)$$

mit

$$\theta_i \sim N(2\theta_{i-1} - \theta_{i-2}, \kappa^{-1}), \quad \text{für } i = 3, \dots, I \quad (2.10)$$

dar;  $\theta_1$  und  $\theta_2$  haben dabei einen unabhängigen, flachen Prior. Die dargestellten Priorverteilungen und Verteilungsannahmen für  $\theta$  gelten analog jeweils auch für  $\phi$  und  $\psi$ .

Für den Präzisionsparameter  $\kappa$  des Alters (beziehungsweise  $\lambda$  für die Periode und  $\nu$  für die Kohorte) wird eine Gammaverteilung  $\kappa \sim G(a, b)$  mit hoher Streuung angenommen, oftmals mit  $a = 1$  und  $b = 0.001$  oder  $a = b = 0.001$ . Da mit höherer Präzision die Parameter stärker geglättet werden, wird beim RW 2 auch häufig ein noch kleinerer Wert für  $b$  verwendet. Die Präzisionsparameter  $\kappa$ ,  $\lambda$  und  $\nu$  werden dann aus den Daten geschätzt.

Das Modell mit RW 1 besitzt einen Vorteil im Gegensatz zum RW 2 Modell: Es löst das in (2.5) beziehungsweise (2.6) beschriebene Identifikationsproblem, indem es jeweils die Transformation mit den kleinsten quadrierten ersten Differenzen auswählt. Somit werden die Parameter, die die APC-Effekte so konstant wie möglich halten, bevorzugt. Folglich können die Effekte im RW1 Modell identifiziert werden. Es gilt aber nach wie vor die Beschränkung, dass ausschließlich nicht-lineare Trends interpretiert werden können. Die APC-Parameter für das RW 2 Modell sind hingegen einzeln nicht identifizierbar und damit auch nicht interpretierbar. Das stellt für die Modellierung der Sterbewahrscheinlichkeiten  $\pi_{ij}$  und die Sterbefälle  $y_{ij}$  allerdings kein Problem dar, da sie nach wie vor eindeutig bestimmt werden können.

## 2.3 Zusätzliche Heterogenität

Das bisher betrachtete bayesianische APC-Modell aus (2.3) kann um den zusätzlichen Heterogenitätsparameter  $z_{ij}$  erweitert werden. Durch die Aufnahme von  $z_{ij}$  kann unstrukturierte, globale Heterogenität, die sich nicht durch die Alters-, Perioden- oder Kohorteneffekte erklären lässt, mit in die Modellierung aufgenommen werden, wie

beispielsweise Überdispersion beziehungsweise Einflüsse durch unbeobachtete Kovariablen für die einzelnen Beobachtungen. Das erweiterte Modell besitzt dann die Form:

$$\eta_{ij} = \log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \mu_{ij} + \theta_i + \phi_j + \psi_k + z_{ij} \quad (2.11)$$

Als Prior für  $z_{ij}$  wird eine Normalverteilung

$$z_{ij} \sim N(0, \delta^{-1}) \quad (2.12)$$

verwendet, mit  $\delta \sim G(a_\delta, b_\delta)$  als Prior für den Präzisionsparameter der globalen Heterogenität.

Bei der Anwendung des erweiterten Modells kann es teilweise trotz Aufnahme des globalen Heterogenitätsparameters zu sehr rauen Schätzungen der APC-Parameter kommen. Dafür können Ausreißer in den Daten oder zusätzliche Heterogenität jeweils einzeln in den Effekten von Alter, Periode oder Kohorte Ursachen sein. Das hat zur Folge, dass die Varianz dieser Effekte überschätzt wird, wodurch die Kreditibilitätsintervalle zu breit werden. Dem kann durch Aufnahme von zusätzlichen Heterogenitätsparametern für die einzelnen Zeitskalen entgegengewirkt werden. Im Modell mit globaler Heterogenität (2.11) kann somit auch die Varianz von  $z_{ij}$  verringert werden. Für eine genauere Betrachtung der Theorie zu zusätzlicher Heterogenität siehe Schmid und Held (2007).

## 2.4 Prädiktion

Der Grundgedanke hinter den APC-Modellen ist, anhand der bisherigen Daten möglichst gute Prognosen über die zukünftigen Sterbewahrscheinlichkeiten  $\pi_{ij}$  in Abhängigkeit von Alter, Periode und Kohorte zu treffen. Dafür werden die Perioden- und Kohorteneffekte für Zeitpunkte in der Zukunft extrapoliert. Da für RW 1 ein konstanter Trend über die Zeit angenommen wird und für RW 2 ein linearer Trend, gilt für die Fortsetzung des Periodenparameters

$$\begin{aligned} \phi_{J+1} &\sim N(\phi_J, \lambda^{-1}) && \text{für RW1} \\ \phi_{J+1} &\sim N(2\phi_J - \phi_{J-1}, \lambda^{-1}) && \text{für RW2.} \end{aligned} \quad (2.13)$$

Für den Kohortenparameter gilt analog

$$\begin{aligned} \psi_{K+1} &\sim N(\psi_K, \nu^{-1}) && \text{für RW1,} \\ \phi_{K+1} &\sim N(2\phi_K - \phi_{K-1}, \nu^{-1}) && \text{für RW2.} \end{aligned} \quad (2.14)$$

Der Altersparameter  $\theta_i$  könnte theoretisch ebenfalls extrapoliert werden. Da aber meistens alle für die Analyse interessierenden Altersgruppen bereits in den Daten gegeben sind, wird diese Möglichkeit nicht weiter betrachtet.

Die Posteriorverteilungen für die vorhergesagten Wahrscheinlichkeiten jeder Altersgruppe im ersten Schritt  $\pi_{i,J+1}$  ergeben sich schließlich aus der Anwendung von (2.3) beziehungsweise (2.11) als

$$\pi_{i,J+1} = \frac{\exp(\eta_{i,J+1})}{\exp(1 + \eta_{i,J+1})} = \frac{1}{\exp(-\eta_{i,J+1})} \quad (2.15)$$



Die Wahrscheinlichkeiten zu späteren Zeitpunkten  $\pi_{i,J+t}$  können analog dazu durch den zugehörigen linearen Prädiktor  $\eta_{i,J+t}$  geschätzt werden. Zu beachten ist, dass für die Anzahl zukünftiger Sterbefälle  $y_{i,J+t}$  ergänzend auch die zukünftige Bevölkerung nach Altersgruppen gegeben sein muss. Ist das nicht der Fall und die konkrete Anzahl der Sterbefälle ist trotzdem von Interesse, muss die künftige Bevölkerungsverteilung extra vorhergesagt werden.

## 2.5 Methoden zur Parameterschätzung

Alle folgenden Auswertungen und somit auch Parameterschätzungen wurden mit dem R-Package *bamp* erstellt. Als Verfahren wird hierbei auf zufallsbasierte Markov-Chain-Monte-Carlo (MCMC) -Verfahren zurückgegriffen. Dabei werden durch Sampling über Zufallsstichproben die einzelnen APC-Parameter geschätzt. Das MCMC-Sampling wird genauer bei Carter und Kohn (1994) und Frühwirth-Schnatter (1994) diskutiert. Ergänzende Verfahren, wie das Nutzen von univariaten Metropolis-Steps für  $\eta_{ij}$  und Gibbs-Steps für die Präzisionsparameter  $\kappa, \lambda, \nu$  und  $\delta$ , finden sich unter anderem bei Besag u. a. (1995). Genauere Informationen zur praktischen Anwendung in *bamp* werden von Schmid und Held (2007) und Knorr-Held und Rainer (2001) beschrieben.

Eine Alternative zu den hier verwendeten MCMC-Verfahren stellt die numerische Approximation durch integrated-nested-Laplace-Approximationen (INLA) dar. Eine aktuelle Anwendung von INLA zur Vorhersage mit APC-Modellen ist bei Riebler und Held (2017) zu finden.

# Kapitel 3

## Modellierung von Todesfällen durch Krebs im Verdauungstrakt

Im folgenden Kapitel werden die bisher besprochenen Modellierungsmöglichkeiten der APC-Modelle auf den Datensatz der Todesfälle durch Krebs im Verdauungstrakt in Deutschland angewendet und evaluiert. Dazu werden unterschiedliche Modelle mit Random Walk erster Ordnung oder Random Walk zweiter Ordnung jeweils mit und ohne Überdispersionsannahme erstellt, deren Parameterschätzer interpretiert und anhand von Gütekriterien verglichen werden.

### 3.1 Beschreibung der Daten

Die Bevölkerungs- und Sterbezahlen stehen jährlich von 1998 bis 2017, also  $J = 20$  Jahre lang, unterteilt nach Geschlecht zur Verfügung. Die jeweiligen Altersgruppen liegen in 5 Jahresintervallen von 25-29 bis 85-89 Jahren vor. Das entspricht  $I = 13$  unterschiedlichen Altersgruppen. Daraus ergeben sich  $K = 5 \cdot (I - 1) + J = 80$  verschiedene Kohortengruppen mit Geburtsjahren von 1911 bis 1991 als jeweiliges Zentrum. Die kommenden Analysen werden jeweils für Frauen und Männer durchgeführt.

In Abbildung 3.1 werden die zeitlichen Verläufe der Todesfälle und Sterberaten durch Krebs nach Altersgruppen dargestellt. Bei Betrachtung der absoluten Anzahlen der Todesfälle lässt sich erkennen, dass tendenziell mit ansteigendem Alter mehr Leute an Karzinomen im Verdauungstrakt sterben, was bei Frauen noch ein wenig stärker zu erkennen ist als bei Männern. Dieses Muster ist aber nicht eindeutig, da in beiden Fällen die älteste Altersgruppe mit 85-89 Jahren nicht am stärksten ausgeprägt ist. Eindeutige Trends der totalen Anzahlen sind insgesamt nicht zu erkennen. Bei den Männern lässt sich beobachten, dass die Todesfälle der Gruppen älter als 74 Jahre in den letzten Jahren angestiegen sind. Bei den restlichen Gruppen der Männern sowie bei allen Altersklassen der Frauen lassen sich eher konstante Verläufe mit unstrukturierten Abweichungen vermuten. Insgesamt gibt es bei den Männern auch eine höhere Anzahl an Todesfällen als bei den Frauen.

Da die Anzahl der jeweiligen Sterbefälle und deren Verlauf aber natürlich stark

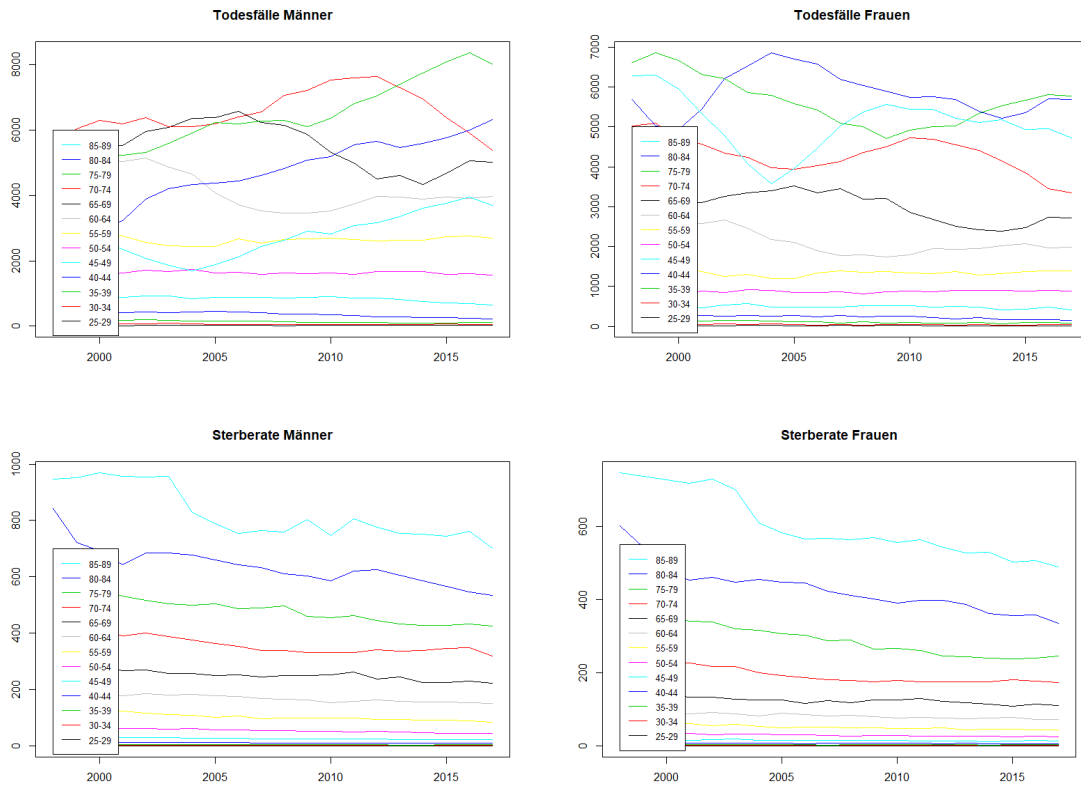


Abbildung 3.1: Todesfälle je Altersgruppe (oben) und Sterblichkeitsrate als Todesfälle je 100 000 je Altersgruppe (unten) nach Geschlecht.

von der Verteilung der Bevölkerung nach den Altersklassen zu den jeweiligen Zeitpunkten abhängt, wird ebenfalls der zeitliche Verlauf der einzelnen Sterberaten betrachtet. Die Sterberaten sind vor allem auch bei der Modellierung der absoluten Sterbefälle  $y_{ij}$  von Bedeutung, da hierfür die benötigten Chancen  $\eta_{ij}$  beziehungsweise Wahrscheinlichkeiten  $\pi_{ij}$  (siehe Gleichungen (2.3) beziehungsweise (2.11)) verwendet werden. Hier kann man eindeutig einen Anstieg der Mortalität mit steigendem Alter erkennen, der über alle Zeitpunkte und für beide Geschlechter vorhanden ist. Zudem lässt sich bei Männern und Frauen ein Absinken der Todesfälle je 100 000 für die einzelnen (vor allem älteren) Altersklassen in den letzten 20 Jahren erkennen. Insgesamt sind die Sterberaten für die Männer jeweils trotzdem noch höher als für Frauen, wobei in beiden Fällen die Sterberaten für jüngere Menschen bis ca. 50 Jahren verhältnismäßig sehr gering sind.

## 3.2 Anwendung von APC-Modellen

Aus der Betrachtung der Daten würden sich bisher zwei monotone Trends vermuten lassen: Erstens ein Ansteigen der Sterbewahrscheinlichkeit mit ansteigendem Alter und zweitens ein Sinken der Sterbewahrscheinlichkeit im zeitlichen Verlauf. In der

bisherigen Darstellung konnte der Einfluss der Kohortenjahrgänge allerdings noch nicht betrachtet werden. Um die Daten genauer auf die drei Einflüsse durch Alter, Periode und Kohorte zu untersuchen und Vorhersagen der Sterbewahrscheinlichkeiten für die Zukunft zu treffen, werden bayesianische APC-Modelle auf den Datensatz angewendet.

Dazu werden die Modelle anfangs auf den gesamten Datensatz gefittet, um die Struktur der gegebenen Daten zu analysieren und die Güte für Interpolation zu vergleichen. Orientiert an bisheriger Fachliteratur (Schmid und Held, 2004), (Knorr-Held und Rainer, 2001) wird innerhalb eines Modells für die drei APC-Parameter jeweils ein Random Walk derselben Ordnung angenommen. Für die beiden Möglichkeiten RW 1 und RW 2 wird jeweils ein Modell ohne globalen Heterogenitätsparameter (2.3) und eines mit globalem Heterogenitätsparameter (2.11) geschätzt. Wie in Kapitel 2.2 allerdings schon erläutert, können die einzelnen Parameter für Alter, Periode und Kohorte für das RW 2 Modell nicht identifiziert werden. Deswegen werden nur die APC-Effekte der RW 1 Modelle graphisch dargestellt, in Abbildung 3.2 ohne Überdispersion und in Abbildung 3.3 mit Überdispersion, jeweils nach Geschlecht getrennt. Für die einzelnen Effekte gilt jedoch nach wie vor, dass aufgrund der möglichen linearen Transformationen der Parameter in den Gleichungen (2.5) beziehungsweise (2.6) ausschließlich nicht-lineare Trends näher interpretiert werden können.

Vergleicht man die APC-Effekte aus den Modellen ohne Aufnahme des globalen Heterogenitätsparameters aus Abbildung 3.2 mit ihren Pendanten mit dessen Aufnahme aus Abbildung 3.3, lassen sich für zweitens, vor allem zu sehen im Kohorteneffekt, glattere Kurven für die Schätzer beobachten. Das entspricht unseren Erwartungen, da durch die Aufnahme des Zusatzparameters  $z_{ij}$  die Varianz der Schätzer für Alter, Periode und Kohorte verringert werden soll. Über die Güte lässt sich keine direkte Aussage treffen, die 90% Kredibilitätsintervalle scheinen sich zwischen beiden Ansätzen nicht groß zu unterscheiden. Da auch wie erwartet keine sichtliche Veränderung im nicht-linearen Verlauf zu sehen ist, können die Effekte beider Modelle gleich interpretiert werden. Es fällt ebenfalls auf, dass sich die Schätzer für Männer und Frauen nicht sehr stark voneinander unterscheiden.

Für das Alter lässt sich in der Parameterschätzung weder bei Frauen noch bei Männern klar eine nicht-lineare Veränderung außer ein leichtes Abflachen der Steigung erkennen. Auch bei den Graphen für die Periodeneffekte sind die Abweichungen von einem linearen Trend nicht auffällig, ab dem Jahr 2002 kann für beide Geschlechter eine etwas stärker abnehmende Steigung betrachtet werden. Da der Datensatz mit 20 Jahren Beobachtungszeitraum aber verhältnismäßig klein und damit anfällig für höhere Einflüsse durch Ausreißer ist, sollte so eine geringe Abweichung nicht überinterpretiert werden. Im geschätzten Kohortenparameter ist mehr Variation vorhanden. Es zeigt sich eine stärkeres Abflachen der Steigung um das Geburtsjahr 1943, für Männer wie Frauen. Für das Ausbrechen der meisten der Krebsarten in den Verdauungsorganen werden vor allem Übergewicht als erhöhender Risikofaktor und Bewegung als senkender Faktor angenommen (Stiftung deutsche Krebshilfe, 2018). Eine theoretisch mögliche Interpretation der abnehmenden Sterbewahrscheinlich-

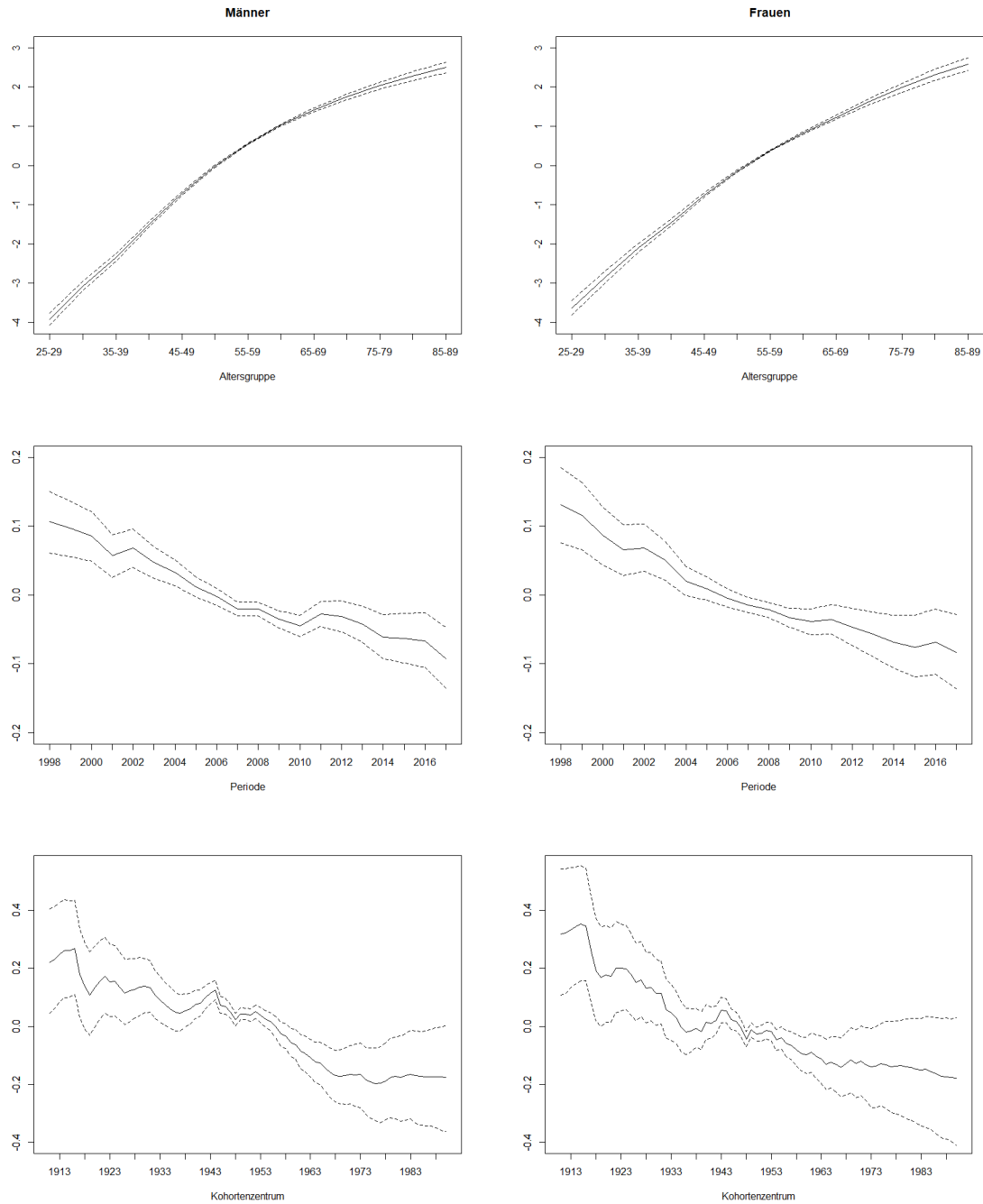


Abbildung 3.2: Posterior geschätzte RW 1 Effekte für Alter, Periode und Kohorte mit Median und 90% Kredititätsintervall im Modell ohne Überdispersion.

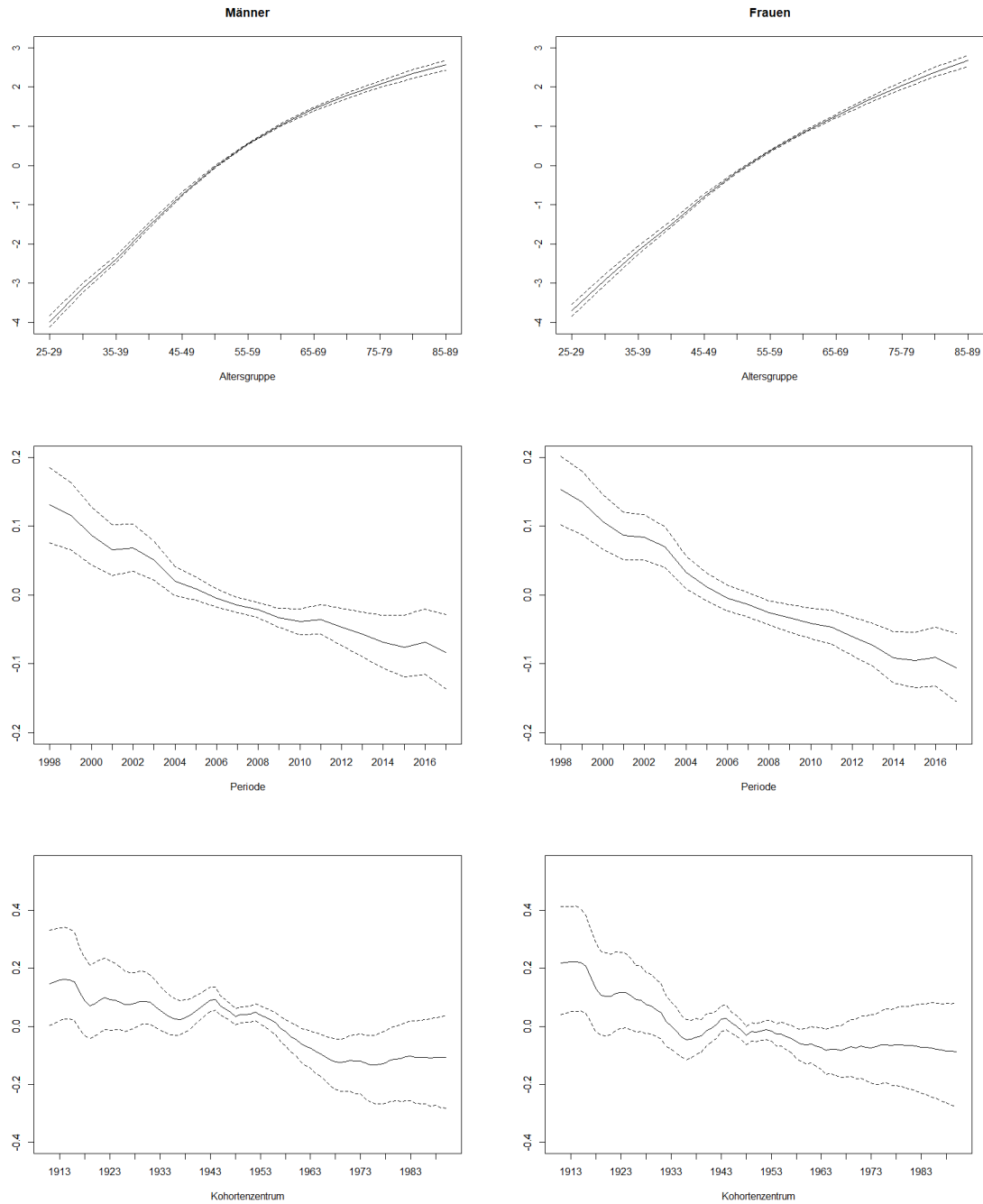


Abbildung 3.3: Posterior geschätzte RW 1 Effekte für Alter, Periode und Kohorte mit Median und 90% Kredititätsintervall im Modell mit Überdispersion.

keiten dieser Kohorten bei Gleichbleiben anderer Bedingungen wäre, dass es durch Nahrungsmittelknappheit nach dem zweiten Weltkrieg und dem Wiedererrichten der durch den Krieg zerstörten Städte unter den jungen Leuten eine niedrigere Quote an Übergewichtigen und mehr körperliche Betätigung gab. Das könnte wiederum die Wahrscheinlichkeit für einen späteren Tod durch Krebs im Verdauungstrakt dieser Generation gesenkt haben. Das sind allerdings nur spekulative Aussagen.

Die Wichtigkeit der einzelnen Schätzer für die Modelle ist an der Spannweite der Graphen erkennbar. In allen vier Modellen ist der Abstand zwischen Maximum und Minimum für den Alterseffekt eindeutig am größten. Dementsprechend hat er den größten Einfluss auf die Modellierung und ist am wichtigsten. Die Spannweite der Perioden- und Kohorteneffekte sind jeweils ungefähr gleich groß und haben somit beide einen ähnlich starken Einfluss. Für eine genauere Betrachtung im Hinblick auf die Stärke der Einflüsse können die Posteriorverteilungen der Präzisionsparameter  $(\kappa, \lambda, \nu)$  der geschätzten Effekte verwendet werden. Dabei sprechen kleine Werte der Schätzer für einen hohen Einfluss und umgekehrt.

### 3.3 Modellevaluierung

Bisher konnten nur einzelne Parameter qualitativ interpretiert werden und das auch nur für die RW 1 Modelle. Dabei ist vor allem von Interesse, wie gut die Anpassung der einzelnen Modelle auf den Datensatz ist. Um die Modellgüte zu bestimmen, wurde jeweils die Posterior-Devianz ermittelt. Diese ergibt sich als

$$D = 2 \cdot \sum_{ij} (l(\hat{y}_{ij}) - l(y_{ij})), \quad (3.1)$$

wobei  $l(\hat{y}_{ij})$  die logarithmierte Likelihood der einzelnen Fälle im Modell ist und  $l(y_{ij})$  die logarithmierte Likelihood der beobachteten Fälle. Da die Devianz selbst eine Posteriorverteilung darstellt, wird für die Evaluierung ihr Posterior-Mittelwert  $\bar{D}$  als Devianz betrachtet. Dabei gilt: Je größer die Werte, desto schlechter die Anpassung des Modells. Wie in den meisten Modellierungen wird auch hier mit zusätzlicher Parameteranzahl die Anpassung des Modells verbessert. Das heißt in diesem Fall, dass durch Aufnahme des globalen Heterogenitätsparameters oder zusätzlicher Parameter für Heterogenität auf den einzelnen Zeitskalen kleinere und damit bessere Devianzen auftreten als ohne deren Aufnahme. (Schmid und Held, 2007)

Um die Modellgüte trotzdem untereinander vergleichen zu können, wird das von Spiegelhalter u. a. (2002) beschriebene Devianz-Informationskriterium (DIC) verwendet. Dabei wird die Devianz mit der effektiven Anzahl an Parametern  $p_D$  bestraft. Diese wird berechnet durch

$$p_D = \bar{D} - D(\bar{\hat{y}}_{ij}), \quad (3.2)$$

wobei  $D(\bar{\hat{y}}_{ij})$  sozusagen den Abstand der echten Werte zu einem Punktschätzer, nämlich dem Mittelwert der geschätzten Posterior der Fälle  $\hat{y}_{ij}$ , darstellt. Das DIC als Kriterium zum Vergleich der Modellgüte ergibt sich damit als

$$DIC = \bar{D} + p_D. \quad (3.3)$$

Auch hier stehen niedrigere Werte für eine bessere Anpassung des Modells an die Daten.

In Tabelle 3.1 sind der Posterior-Median der Devianz, die effektive Anzahl an Parametern  $p_D$  und das DIC für alle Modelle dargestellt. Bezüglich der Devianz besitzen bei den Modellen ohne Überdispersionsannahme die RW 1 Modelle im Vergleich zu den RW 2 Modellen bei Männern und Frauen die kleineren und damit besser auf die Daten angepassten Werte. Die Modelle mit Überdispersion zeigen erwartungsgemäß die niedrigste Devianz durch sehr genaue Anpassung an die Daten auf und unterscheiden sich daher auch nur marginal innerhalb der RW 1 und RW 2 Modellierungen. In der Bewertung durch das DIC stellen sich allerdings sowohl für Random Walk erster Ordnung als auch für Random Walk zweiter Ordnung die Modelle ohne Aufnahme eines globalen Heterogenitätsparameters bei beiden Geschlechtern als geeigneter heraus. Das liegt an der größeren Anzahl der effektiv geschätzten Parameter  $p_D$  in den Modellen mit Überdispersion im Vergleich zu denen ohne. Der RW 1 zeigt im Vergleich mit RW 2 bei den Modellen ohne Überdispersion im DIC noch leicht bessere Werte auf. Die Differenzen der DICs mit 4.2 bei den Männern und 5.9 bei den Frauen sind dabei aber so niedrig, dass nicht direkt von einer besseren Anpassung der RW 1 Modelle ausgegangen werden kann.

Männer			
Überdispersion	Devianz	$p_D$	DIC
Random Walk 1			
ohne	271.9	71.2	343.0
mit	229.0	167.4	396.3
Random Walk 2			
ohne	289.1	58.1	347.2
mit	229.8	164.3	394.0
Frauen			
Überdispersion	Devianz	$p_D$	DIC
Random Walk 1			
ohne	286.1	70.8	357.0
mit	235.1	160.5	395.6
Random Walk 2			
ohne	306.0	56.9	362.9
mit	235.9	156.0	392.0

Tabelle 3.1: Posterior-Median der Devianz, effektive Anzahl an Parametern und DIC für alle Modelle.

### 3.4 Einbinden von Kovariablen in APC-Modelle

In Kapitel 3.2 wurden bei der Interpretation des Kohorteneffektes schon die möglichen Zusammenhänge von Übergewicht mit der Sterblichkeit durch Krebs in den Verdau-



ungsorganen angesprochen (Stiftung deutsche Krebshilfe, 2018). Von Boeing u. a. (1991) wurden speziell für Magenkrebs weitere mögliche Ursachen diskutiert. Durch die Zusammenfassung mehrerer Krebsarten, wie Magenkrebs, Darmkrebs und weiterer Arten, häufen sich auch die Möglichkeiten von verschiedenen Einflüssen für die in dieser Arbeit modellierten Sterbewahrscheinlichkeiten. Deswegen wäre eine Verbesserung der Modellierungen auch bei der Verfügbarkeit von sehr gut dokumentierten Daten zu Übergewicht, Bewegung oder anderen theoretischen Risikofaktoren nicht so sicher anzunehmen, wie es bei Todesfällen durch andere Krankheitsarten der Fall ist. Ein Beispiel hierfür ist der Zusammenhang von Zigarettenkonsum und der Sterbewahrscheinlichkeit durch Lungenkrebs, der ebenfalls durch APC-Modelle mit dem Einbinden des Tabakkonsums als Kovariable in Knorr-Held und Rainer (2001) genauer untersucht wird. In unserem Beispiel wird auf die Vermutung des Einflusses des Übergewichts im zeitlichen Verlauf auf die Mortalität genauer eingegangen. Stellvertretend dafür wird in dieser Analyse der durchschnittliche Body-Mass-Index (BMI) als Indikator verwendet, bei dem ein höherer Wert eher für Übergewicht steht (für genauere Informationen siehe WHO (2000)). Der BMI in Deutschland wird von NCD Risk Factor Collaboration (2019) für die Jahre 1975 bis 2016 als altersstandardisierter durchschnittlicher BMI getrennt nach Geschlecht, einmal für Erwachsene und einmal für Kinder von 5 bis 19 Jahren, zur Verfügung gestellt.

Eine Möglichkeit ist, den Periodeneffekt durch den Einfluss der Kovariable durchschnittlicher BMI von Erwachsenen zu ersetzen. Das Modell (2.11) kann wie folgt abgeändert werden:

$$\eta_{ij} = \mu + \theta_i + \beta_j x_{j-L} + \psi_k + z_{ij}, \quad (3.4)$$

wobei  $z_{ij}$  optional auch wieder weggelassen werden kann. Für den Regressionsparameter  $\beta_j$  kann analog zum ersetzten  $\theta_j$  wieder eine RW 1 oder eine RW 2 Priorverteilung angenommen werden (Schmid und Held, 2007). Die zeitliche Verschiebung um  $L$  Jahre in  $x_{j-L}$  ist sinnvoll, da sich häufig die Auswirkung der Kovariable auf die Sterbewahrscheinlichkeit erst nach einer gewissen Zeit zeigt. In der bereits erwähnten Analyse von Knorr-Held und Rainer (2001) zum Lungenkrebs wurde ein Zeitraum von ungefähr  $L = 20$  Jahren angenommen, bis die Auswirkungen durch das Rauchen eintreten. Der BMI an sich hat vermutlich keinen direkten Zusammenhang zur Sterblichkeit durch Krebs im Verdauungstrakt, sondern ist viel mehr das spätere Ergebnis längerfristiger ungesunder Ernährungsweisen, wie beispielsweise hoher Zucker- und Fettkonsum, mangelnde Bewegung und erhöhter Alkoholkonsum, die eher direkten Einfluss auf Krebs im Verdauungstrakt haben können (Stiftung deutsche Krebshilfe, 2018). Dadurch wird das Festlegen einer Verschiebung erschwert beziehungsweise ist nicht eindeutig, ob sie überhaupt nötig ist. Da die Daten zum BMI nur bis zum Jahr 2016 gehen und keine genaueren Angaben zu einer sinnvollen zeitlichen Verschiebung bekannt sind, wird naiv eine Verschiebung von  $L = 1$  Jahr angenommen.

In Abbildung 3.4 ist der Verlauf des durchschnittlichen BMIs über die Jahre für Männer und Frauen dargestellt. Da der Verlauf bei den Erwachsenen annähernd linear steigt, wird nicht davon ausgegangen, dass er die schwachen nicht-linearen Trends in den oben beschriebenen Periodeneffekten der RW1 Modelle erklären kann.

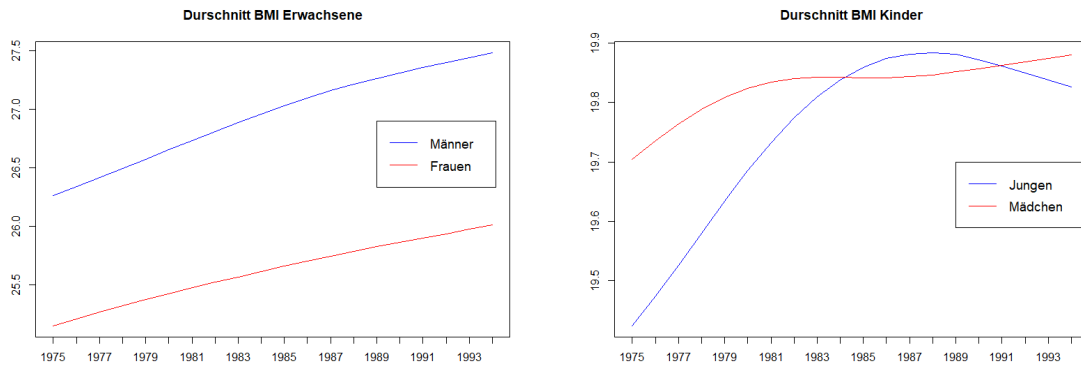


Abbildung 3.4: Altersstandardisierter durchschnittlicher BMI für Erwachsene und Kinder zwischen 5 und 19 Jahren in Deutschland.

Zur Überprüfung, ob die Modellgüte durch die Aufnahme der Kovariablen verbessert werden kann, wurden die oben verwendeten Modelle jeweils mit Aufnahme des BMIs als Periodeneffekt mit dem in den restlichen Effekten verwendeten Prior erneut berechnet. Hierbei ließen sich nur minimalste Abweichungen in Devianz und DIC im Vergleich zu ihren Pendanten mit Random Walk gleicher Ordnung als Prior für die APC-Parameter erkennen, die nicht auf eine Verbesserung der Modellierung hinweisen. Auch für die jeweiligen geschätzten Präzisionsparameter ergaben sich nur geringe Unterschiede.

Eine andere Annahme könnte sein, dass sich ein hoher BMI in der Kindheit negativ auf die spätere Sterbewahrscheinlichkeit durch Krebs im Verdauungstrakt auswirkt. Die inhaltliche Überlegung dahinter ist, dass sich oft ungesunde Lebensweisen in der Kindheit etablieren, die sich durch einen frühen hohen BMI aufzeigen können und im Erwachsenenalter fortgeführt werden, was zu Krebs führen könnte. In Abbildung 3.4 ist ebenfalls der Verlauf des durchschnittlichen BMIs für Kinder gegeben. Dabei kann beobachtet werden, dass der Trend vor allem für Jungen in den letzten Jahren nicht mehr annähernd linear steigend ist. Es könnte über ein Ersetzen des Periodenparameters  $\psi_k$  nachgedacht werden. Nimmt man das arithmetische Mittel der 5-19 Jährigen, also ein Alter von 12 Jahren, stellvertretend für den durchschnittlichen altersstandardisierten Wert des BMIs bei Kindern an und verwendet den Wert zu diesem Alter gleichzeitig als Einfluss, ist die Kovariable dementsprechend für die Geburtsjahre ab 12 Jahren vor der ersten gegebenen Erhebung 1975, also ab 1963, verfügbar. Da somit nur ungefähr die Hälfte der Beobachtungen des Datensatzes für eine weitere Modellierung verwendet werden könnte, wird nicht von einer Modellverbesserung durch Aufnahme dieser Kovariable ausgegangen.

### 3.5 Weitere Möglichkeiten zur Modellierung

Bisher wurden immer innerhalb eines Modells für die Parameterschätzer jeweils Random Walks der gleichen Ordnung als Prior angenommen. Es besteht aber genauso

die Möglichkeit, innerhalb eines Modells unterschiedliche Annahmen für die Priorverteilungen der Parameter zu treffen. Des Weiteren können neben dem globalen Heterogenitätsparameter auch ein oder zwei Schätzer für die Alters-, Perioden- oder Kohortenparameter aus der Modellierung gelassen und nicht weiter beachtet werden. Außerdem besteht die Möglichkeit, die Modelle durch Aufnahme zusätzlicher Heterogenitätsparameter auf den einzelnen Zeitskalen, wie in Kapitel 2.3 beschrieben, zu erweitern. Eine von vornherein inhaltliche Überlegung zur Wahl der verschiedenen Optionen zur Modellierung ist aufgrund der Abhängigkeiten von Parametern untereinander und einer Vielzahl möglicher Kombinationen schwer. Für unseren Datensatz wurden stichprobenartig weitere Modelle mit den in diesem Kapitel beschriebenen Möglichkeiten durchgeführt. Anhand des DIC konnte jedoch keine Verbesserung der Modellgüte der neuen Modellen gegenüber den Modellen ohne Überdispersion festgestellt werden.

# Kapitel 4

## Retrospektive Prädiktion

Die bisher betrachteten Modelle wurden anhand des gesamten Datensatzes geschätzt. Die Modellgüte für den Vergleich der Modelle untereinander wurde primär über das DIC festgelegt. Dabei wird bei diesem Kriterium allerdings nur die Anpassung auf die bisherigen Daten betrachtet. Somit ist das DIC hauptsächlich für die Interpolation von Daten von Bedeutung. Da jedoch eine gute Modellierung für die beobachteten Bereiche keine Garantie für gute Extrapolation ist, werden für die Genauigkeit von Vorhersagen weitere Kriterien verwendet.

### 4.1 Methodik

Um Aussagen über die Genauigkeit der Prädiktionen in der Vergangenheit zu machen, wurde die Idee aus Schmid und Held (2004) und Knorr-Held und Rainer (2001) adaptiert, Beobachtungen der letzten Jahre zu löschen, diese anschließend vorherzusagen und wahre und geschätzte Werte schließlich zu vergleichen. In unserem Datensatz wurden für die neuen Modellierungen die Daten bis einschließlich des Jahres 2012 verwendet und die nächsten 5 Jahre vorhergesagt. Dafür wird die in Kapitel 2.5 beschriebene Fortsetzung der Schätzungen der Perioden- und Kohorteneffekte angewendet. Als Modellierungsarten wurden erneut die in Kapitel 3.2 dargestellten RW 1 und RW 2 Modelle mit und ohne globalen Heterogenitätsparameter betrachtet, da auch die Modelle mit Aufnahme der zusätzlichen Kovariablen keine starke Abweichungen in Güte und Schätzern zeigten. Zudem kann so für den Datensatz untersucht werden, ob die Modelle mit der besten Interpolation auch die besten extrapolierten Werte für die Sterbewahrscheinlichkeit durch Krebs in den Verdauungsorganen für die Jahre 2013 bis 2017 liefern.

Um die Güte der Vorhersagen zu bewerten, werden Kredibilitätsintervalle zu den Niveaus von 50%, 80% und 90% für die Anzahlen der Sterbefälle  $y_{ij}$  berechnet. Anschließend wird für die vorhergesagten Fälle mit  $i = 1, \dots, 13$  und  $j = 16, \dots, 20$  der Anteil der innerhalb liegenden tatsächlich beobachteten Werte bestimmt, also für insgesamt  $13 \cdot 5 = 65$  Beobachtungen.

Eine weitere Möglichkeit wäre die Bestimmung der Devianz der Vorhersagen  $D^P$ ,

welche sich als

$$D^P = 2 \cdot \sum_{i,j} (l(\hat{y}_{ij}) - l(y_{ij})) \quad (4.1)$$

darstellen lässt, wobei  $\hat{y}_{ij}$  die vorhergesagten Fälle sind und somit für den Periodenindex  $j$  in unserem Fall wieder  $j = 16, \dots, 20$  verwendet werden würde (Schmid und Held, 2007).

Die bisherigen Methoden bewerten die Güte der Modelle für Interpolation sowie Extrapolation hauptsächlich auf Basis geschätzter Posteriorverteilungen. Aus Gründen der Interpretierbarkeit und zur Vereinfachung weiterer Schätzungen sind oft konkrete Punktschätzer als Werte für die Vorhersage interessant. Dafür wird in den graphischen Darstellungen der Fachliteratur meistens der Median der Posteriorverteilungen der Fälle beziehungsweise der Wahrscheinlichkeiten betrachtet. Dementsprechend wird in dieser Arbeit als weiteres Kriterium für die bisherige Prädiktionsgüte die Devianz des vorhergesagten Posterior-Medians  $D^P(\hat{y}_{med})$  eingeführt. Orientiert an der für das DIC in (3.3) beziehungsweise (3.2) verwendeten Devianz des Posterior-Mittelwerts  $D(\hat{y}_{ij})$  und an Knorr-Held und Rainer (2001) ergibt sich dann

$$D^P(\hat{y}_{med}) = \sum_i \sum_j v_{ij}, \quad (4.2)$$

mit

$$v_{ij} = 2 \cdot y_{ij} \cdot \log \left( \frac{y_{ij}}{y_{ij}^{med}} \right) + 2 \cdot (n_{ij} - y_{ij}) \cdot \log \left( \frac{n_{ij} - y_{ij}}{n_{ij} - y_{ij}^{med}} \right) \quad (4.3)$$

als Maß für den Abstand zum individuellen Posterior-Median  $y_{ij}^{med}$  für die Vorhersagen der Perioden  $j$  in den Altersgruppen  $i$ .

## 4.2 Bewertung der Vorhersageperformance

In Tabelle 4.1 sind die Anteile der beobachteten Fälle innerhalb der vorhergesagten Kredibilitätsintervalle zu den Niveaus von 50%, 80% und 90%, sowie die Devianz des vorhergesagten Medians, für die Modelle dargestellt. Die Anteile sind für die Modelle mit Aufnahme von globaler Heterogenität wie erwartet ceteris paribus höher als ohne. Da bei den Modellen ohne Überdispersion allerdings schon meist ein höherer Anteil an beobachteten Fällen, als durch das Kredibilitätsniveau vermutet, innerhalb der Intervalle liegt, wird die Prädiktionsgüte durch die Annahme von Überdispersion eher verschlechtert. Auch in der Devianz der Mediane zeigen sich eindeutig größere Abweichungen des Punktschätzers zu den wahren Werten. Im Vergleich von RW 1 und RW 2 Modellen fällt auf, dass das RW 2 Modell ebenfalls zu breite Kredibilitätsintervalle für Männer und Frauen schätzt. Die RW 1 Modelle liefern noch die den gewählten Kredibilitätsniveaus am nächsten kommenden Anteile. Interessanterweise liefern die RW 2 Modelle unter Verwendung des Medians als Punktschätzer genauere Vorhersagen für konkrete Werte, bei Männern sowie bei Frauen, was durch

Männer				
	Kreditabilitätsniveau			Devianz
Überdispersion	50%	80%	90%	$D^P(\hat{y}_{med})$
	Random Walk 1			
ohne	0.42	0.78	0.94	355.6
mit	0.54	0.86	0.98	510.0
	Random Walk 2			
ohne	0.78	0.94	0.98	285.4
mit	0.78	0.94	0.98	501.7
Frauen				
	Kreditabilitätsniveau			Devianz
Überdispersion	50%	80%	90%	$D^P(\hat{y}_{med})$
	Random Walk 1			
ohne	0.45	0.88	0.95	207.5
mit	0.52	0.95	0.98	344.1
	Random Walk 2			
ohne	0.80	0.89	0.95	92.9
mit	0.83	0.97	0.95	101.8

Tabelle 4.1: Anteile beobachteter Fälle innerhalb der vorhergesagten Kreditabilitätsintervalle und Devianz des vorhergesagten Medians.

die kleineren Werte für  $D^P(\hat{y}_{med})$  zu erkennen ist. Anzumerken ist, dass die Anteile der beobachteten Fälle innerhalb der vorhergesagten Kreditabilitätsintervalle, wie beschrieben, aus nur jeweils 60 Beobachtungen bestimmt werden und damit einzelne Ausreißer einen verhältnismäßig eher hohen Einfluss auf die Anteile haben.

In Abbildung 4.1 werden die beobachteten und vorhergesagten Sterbefälle durch Karzinome im Verdauungstrakt mit Median und 90% Kreditabilitätsintervallen für die Modelle ohne Überdispersion gezeigt. Diese sind nach Altersgruppen aggregiert. Aufgrund der bisherigen schlechteren Bewertungen der Modelle mit globalem Heterogenitätsparameter werden fortan nur noch die Modelle ohne Annahme von Überdispersion genauer betrachtet. Hier kann man ebenfalls bei den RW 2 Modellen größere Kreditabilitätsintervalle zum Niveau von 90% im Vergleich zu den RW 1 Modellen sehen. Es kann gesagt werden, dass mit beiden Modellierungsarten die absolute Anzahl an Sterbefällen pro Jahr tendenziell überschätzt wurde, was sich durch die Position der Datenpunkte unterhalb der geschätzten Mediane bemerkbar macht. Unter Betrachtung des Medians als Punktschätzer zeigen sich für beiden Geschlechter im RW 2 Modell kleinere Abweichungen zu den beobachteten Daten, was für bessere Vorhersagen spricht. Unter diesem Aspekt lässt sich vor allem beim RW 2 Modell für Frauen eine äußerst gute Vorhersagegenauigkeit für die Jahre 2012 bis 2017 feststellen.

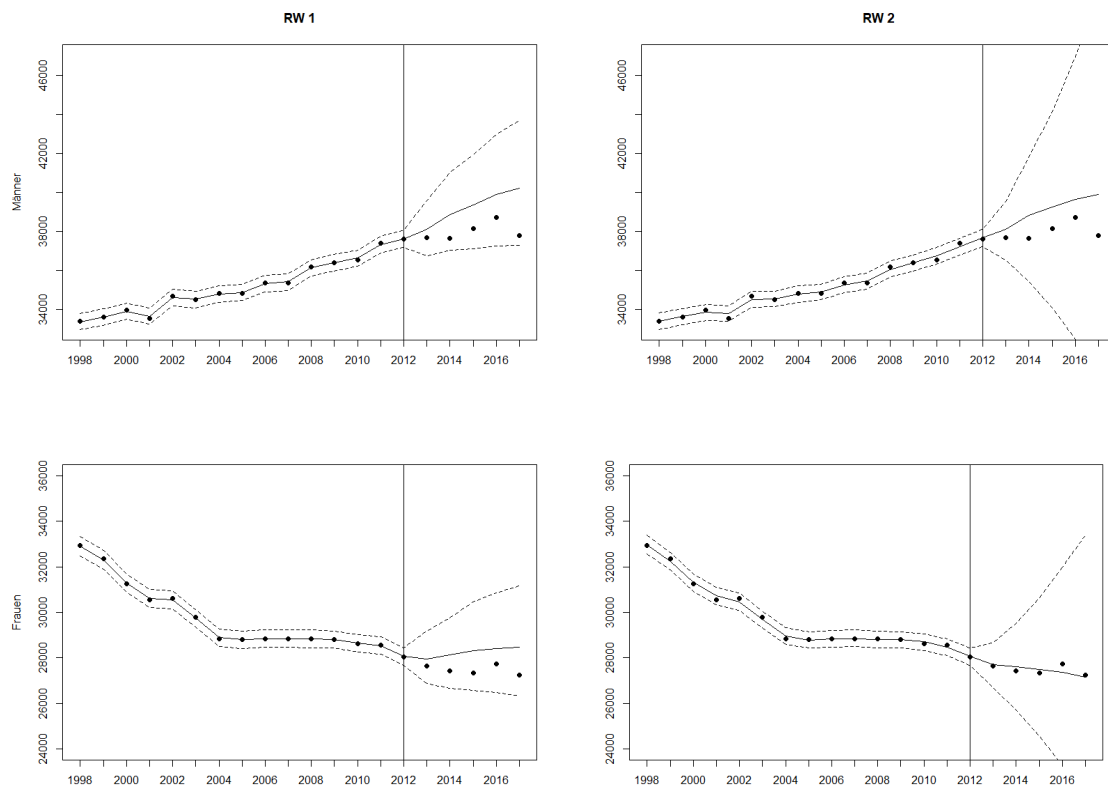


Abbildung 4.1: Beobachtete und ab 2012 vorhergesagte Sterbefälle mit Median und 90% Kreditäritätsintervall aggregiert über alle Altersgruppen; für Männer (oben) und Frauen (unten) mit RW 1(links) und RW 2 (rechts) ohne Überdispersion.

# Kapitel 5

## Vorhersage für zukünftige Jahre

Bisher wurden verschiedene Möglichkeiten für die Modellierung mit Alters-Perioden-Kohorten-Modellen vorgestellt, geschätzt und deren Anpassung auf gegebene Daten evaluiert. Das Ziel dahinter ist, für unseren Datensatz ein möglichst gut geeignetes Modell zur Vorhersage der Sterbewahrscheinlichkeiten beziehungsweise der zukünftigen Todesfälle durch Krebs im Verdauungstrakt für die nächsten Jahren zu finden.

### 5.1 Modellwahl für die Prädiktion

In den vorherigen Kapiteln wurde einerseits die Modellgüte für Interpolation und andererseits die Güte für Vorhersagen der letzten 5 erhobenen Jahresdaten bewertet. Dabei konnte bei beiden Vorgängen festgestellt werden, dass die Aufnahme eines zusätzlichen globalen Heterogenitätsparameters die Modellierung im Falle unserer Daten nicht verbessern konnte. Des Weiteren konnte auch kein Mehrwert durch die Aufnahme des durchschnittlichen BMI als Kovariable für den Periodeneffekt beobachtet werden.

Bei dem Vergleich der Modelle mit RW 1 zu den Fällen mit RW 2 als Prior für die APC-Parameter konnte keine der beiden Modellierungen eindeutig als besser klassifiziert werden. Die RW 1 Modelle zeigten minimalst bessere Werte für das DIC auf. Bei der Überprüfung der Prädiktionsgüte in den letzten Jahren waren für die RW1 Modelle bessere Werte für die Anteile beobachteter Daten innerhalb der vorhergesagten Kreditabilitätsintervalle zu den Niveaus 50%, 80% und 90% zu erkennen, da die RW 2 Modelle jeweils zu breite Intervalle schätzten. Für den Abstand des Medians der vorhergesagten zu den tatsächlichen Werten erzielten allerdings die RW 2 Modelle klar bessere Ergebnisse. Je nachdem, ob der Fokus bei den Vorhersagen auf Posteriorverteilungen oder auf Punktschätzer gelegt wird, können in unserem Fall beide Modellierungen von Interesse sein. Deswegen werden für die kommenden Schätzungen der Prädiktionen für die Jahre 2018 bis 2022 weiterhin sowohl RW 1 als auch RW 2 Modelle angewendet und betrachtet.



## 5.2 Interpretation der Prädiktionen

In den Abbildungen 5.1 bis 5.4 sind die beobachteten und vorhergesagten Sterberaten als Sterbefälle je 100 000 mit Median und 90% Kredibilitätsintervallen dargestellt, jeweils getrennt nach Modell, Geschlecht und Altersgruppe. Hierbei interessieren wir uns nur für die Altersgruppen von 50 Jahren aufwärts, da durch die niedrigen Sterberaten in den jungen Altersgruppen (siehe Abbildung 3.1) nur ungenaue und für die Vorhersage eher unrelevante Schätzungen möglich sind.

Als globaler zeitlicher Trend kann über beide Modellierungen ein Abnehmen der Sterberaten für die meisten Altersgruppen bei beiden Geschlechtern, sowohl in der Vergangenheit, als auch in den Prognosen für die Jahre bis 2022, betrachtet werden. Das gilt vor allem für die Altersgruppen von 50 bis 69 Jahren. Prinzipiell nimmt das RW 2 Modell in unserer Analyse durch das Fortführen von linearen Trends stärker sinkende Raten im Median für die Zukunft an als das RW 1 Modell. Dafür können auch hier wieder breitere 90% Kredibilitätsintervalle und glattere Mediane der Schätzer beim RW 2 Modell festgestellt werden. Für die Altersgruppen ab 70 Jahren lassen sich bei beiden Geschlechtern teilweise stagnierende vorhergesagte Raten beziehungsweise vereinzelte Anstiege in der Prognose der 5 Jahre in den RW 1 Modellen erkennen, während im RW 2 Modell trotzdem noch eher von sinkenden Raten ausgegangen wird. Das liegt an der Annahme der konstanten Trends als Prior für die APC-Effekte im RW 1 Modell, der sich auf die Fortsetzung der Perioden- und Kohorteneffekte auswirkt.

Zusammengefasst: Es werden mit dem RW 2 Modell für die Zukunft tendenziell niedrigere Sterberaten durch Krebs im Verdauungstrakt in den Altersgruppen vorhergesagt als mit dem RW 1 Modell. Zudem machen sich keine sehr starken Unterschiede zwischen den Geschlechtern in den vergangenen und prognostizierten Raten bemerkbar. In den jüngeren Altersgruppen wird künftig von sinkenden Sterberaten durch Karzinome im Verdauungstrakt ausgegangen, in den älteren je nach Modell von stagnierenden bis leicht sinkenden. Um Prognosen für aggregierte totale Anzahlen an Sterbefällen anzugeben, müsste die Bevölkerung und ihre Altersverteilung in den vorhergesagten Jahren 2018 bis 2022 bekannt sein. Alternativ könnte diese extra prognostiziert werden. Durch die Verwendung dieser geschätzten Bevölkerung zur Vorhersage der Anzahl an Sterbefällen in der Zukunft würde aber gleichzeitig weitere Unsicherheit in die Modellierung gebracht werden (Knorr-Held und Rainer, 2001).

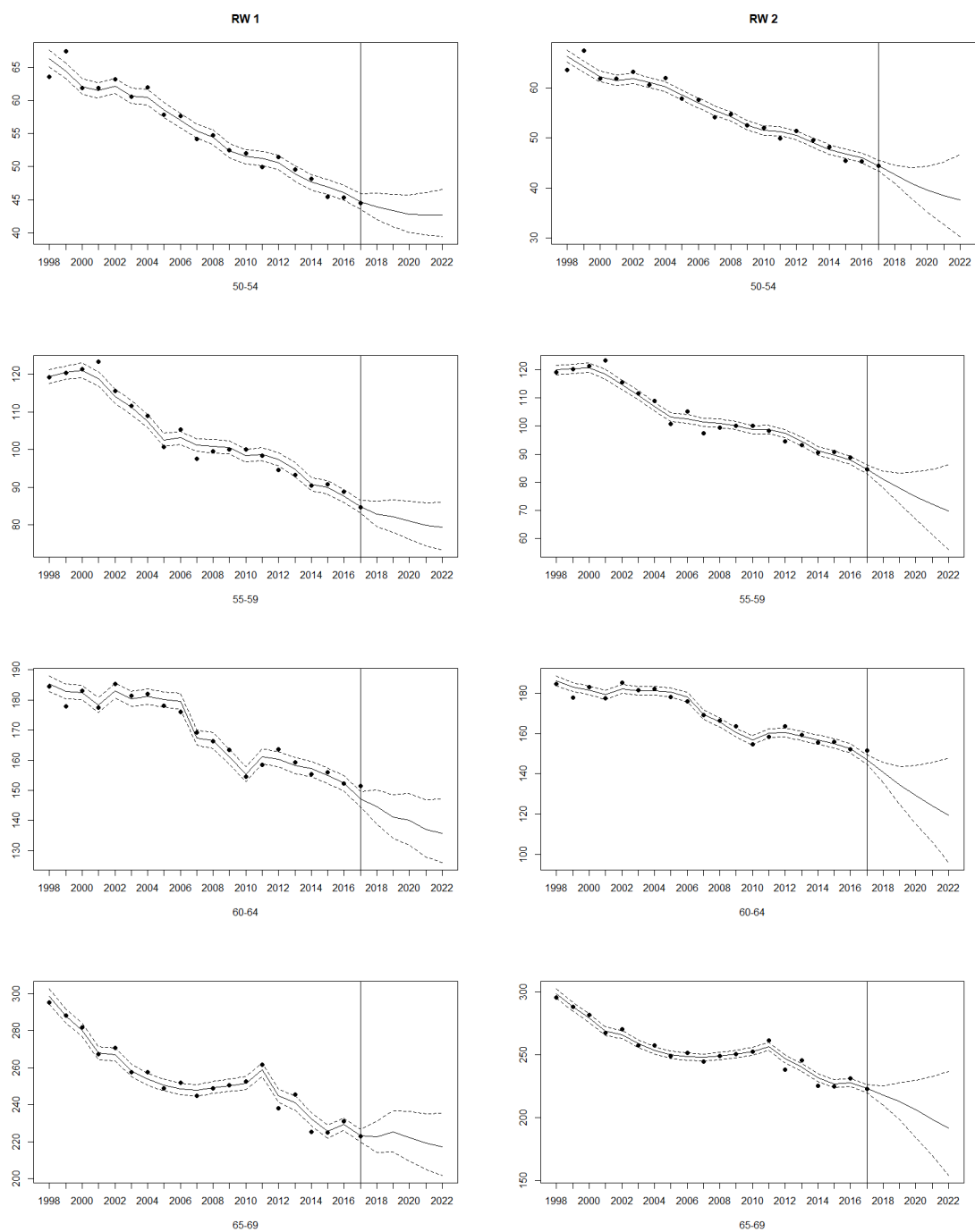


Abbildung 5.1: Beobachtete und vorhergesagte Sterberaten als Sterbefälle je 100 000 mit Median und 90% Kreditätsintervall; für Männer in Altersgruppen 50 bis 69 Jahre mit RW 1(links) und RW 2 (rechts) ohne Überdispersion.

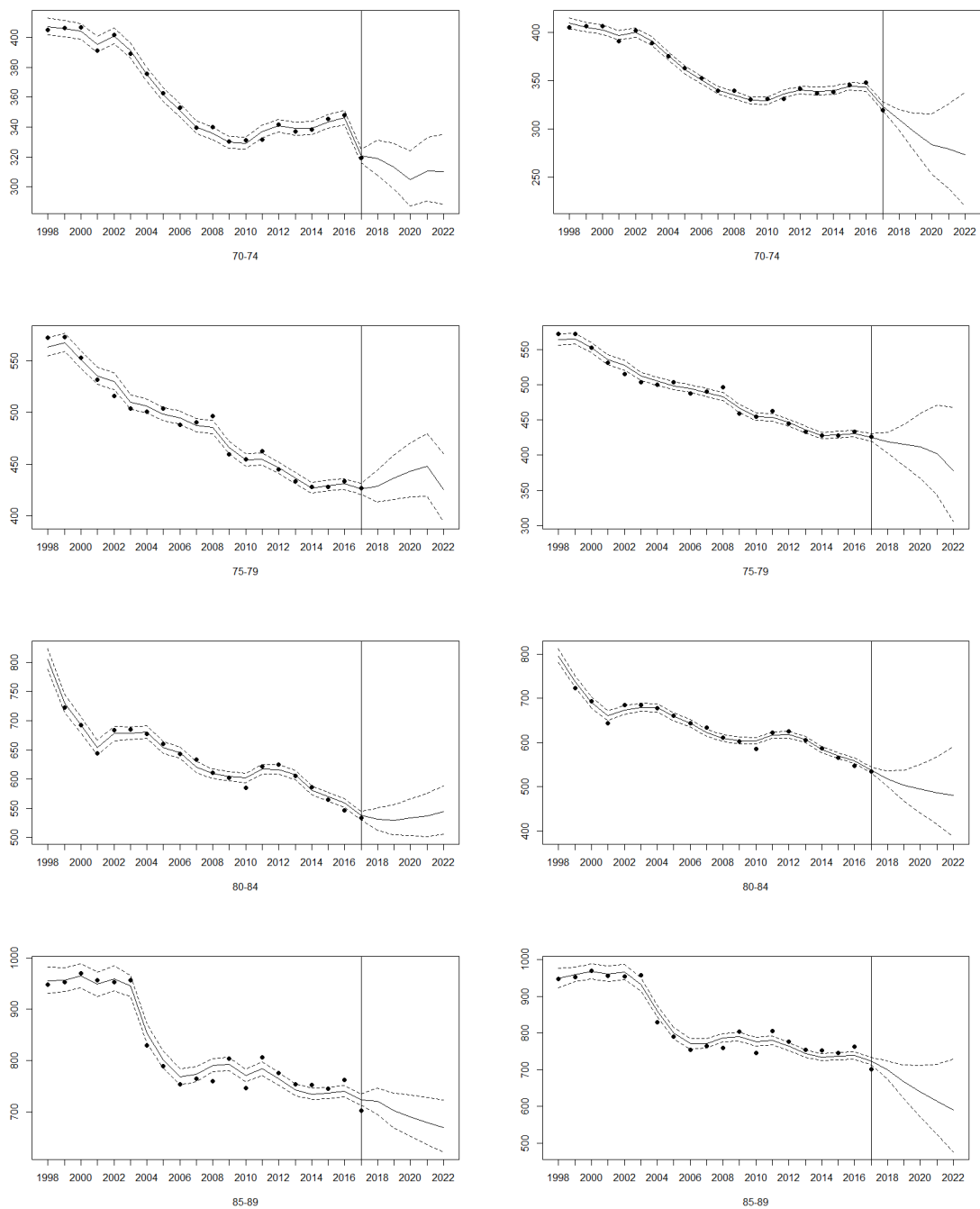


Abbildung 5.2: Beobachtete und vorhergesagte Sterberaten als Sterbefälle je 100 000 mit Median und 90% Kredititätsintervall; für Männer in Altersgruppen 70 bis 89 Jahre mit RW 1(links) und RW 2 (rechts) ohne Überdispersion.

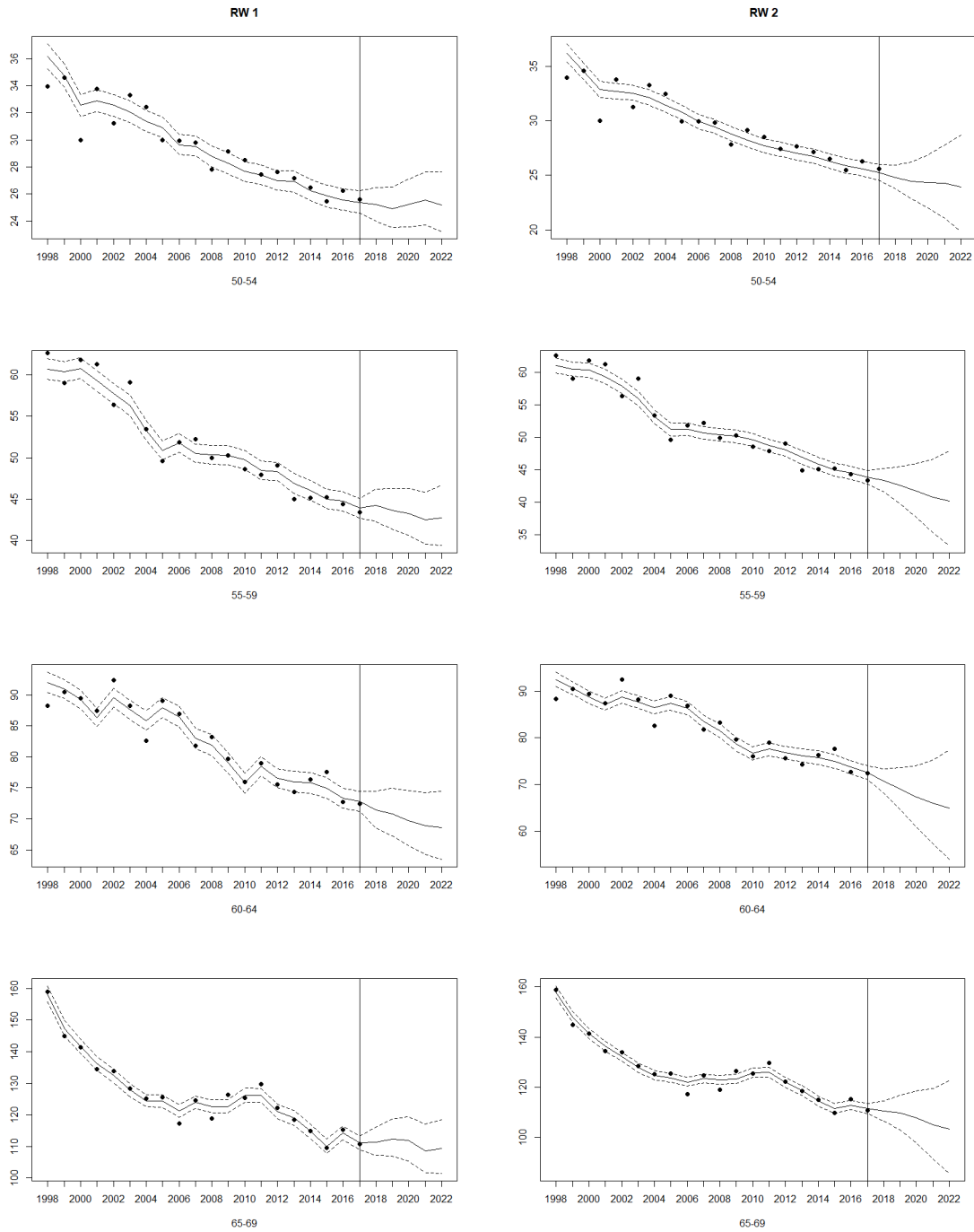


Abbildung 5.3: Beobachtete und vorhergesagte Sterberaten als Sterbefälle je 100 000 mit Median und 90% Kredititätsintervall; für Frauen in Altersgruppen 50 bis 69 Jahre mit RW 1(links) und RW 2 (rechts) ohne Überdispersion.

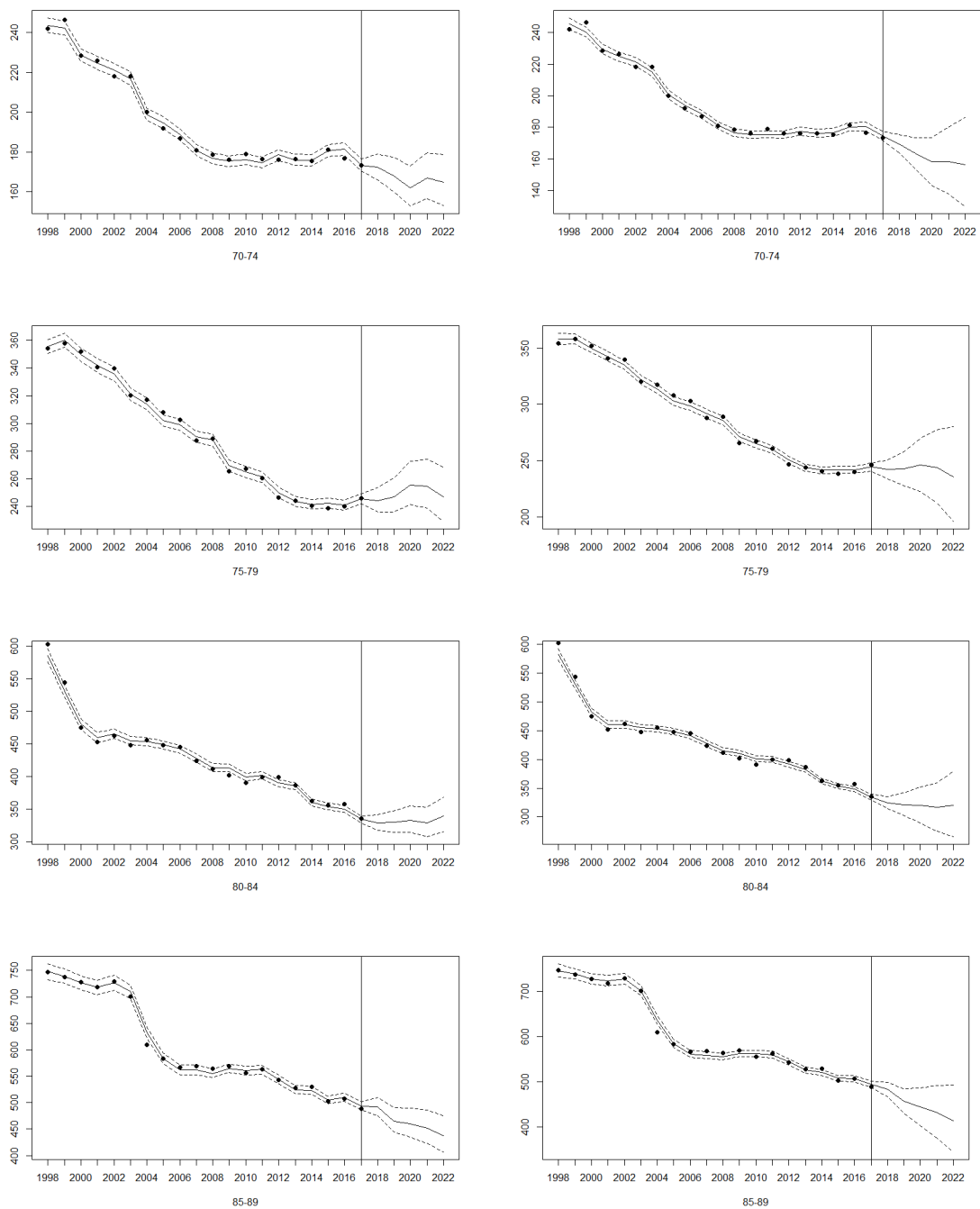


Abbildung 5.4: Beobachtete und vorhergesagte Sterberaten als Sterbefälle je 100 000 mit Median und 90% Kreditsintervall; für Frauen in Altersgruppen 70 bis 89 Jahre mit RW 1(links) und RW 2 (rechts) ohne Überdispersion.

# Kapitel 6

## Konklusion

Ziel der Arbeit war es, verschiedene Möglichkeiten zur Modellierung und Prädiktion durch APC-Modelle darzustellen und diese auf den Datensatz der Sterblichkeit durch Krebs in den Verdauungsorganen anzuwenden, sowie diese miteinander zu vergleichen. Dazu wurden für die Priorverteilungen der Parameter der Alters-, Perioden- und Kohorteneffekte Random Walks erster und zweiter Ordnung betrachtet und die Aufnahme von ergänzenden Parametern für globale Heterogenität und für zusätzliche Heterogenität auf den einzelnen Zeitskalen besprochen. Als weitere Option wurde das Ersetzen der Parameterschätzungen für Perioden- oder Kohorteneffekte durch eine Schätzung über den Einfluss von weiteren Kovariablen vorgestellt.

Für die Prognosen der Sterberaten durch Krebs im Verdauungstrakt in Deutschland konnte auf Basis von Daten der letzten 20 Jahre weder durch die Überdispersionsannahme, noch durch das Aufnehmen des durchschnittlichen BMIs als Ergänzung zum Periodeneffekt, eine Verbesserung der Genauigkeit ausgemacht werden. Prinzipiell ist die Modellerweiterung mit Aufnahme von Kovariablen als Ergänzung zu den geschätzten Einflüssen durch Periode oder Kohorte in anderen Fällen oft eine sinnvolle Methode, um die Prädiktionsgüte zu verbessern und validere Prognosen anzustellen. Das ist vor allem in Themengebieten der Fall, in denen von starken Zusammenhängen zwischen einer Kovariable und der Zielvariable ausgegangen wird. Ein Beispiel dafür ist die bereits erwähnte Analyse von Sterberaten durch Lungenkrebs in Abhängigkeit vom Rauchverhalten von Knorr-Held und Rainer (2001). Ein Vorteil für die Vorhersage durch derart erweiterte Modelle ist, dass durch die besprochene Zeitverschiebung  $L$  der Kovariableneinflüsse für die Periode beziehungsweise durch die Differenz des Alters zum Geburtsjahr für die Kohorte tatsächlich erhobene und nicht etwa geschätzte Werte als Einfluss für zukünftige Sterberaten zur Verfügung stehen. Das bedeutet aber wiederum, dass die Daten in diesem Abstand auch vor den Erhebungszeitpunkten der Bevölkerungszahlen und Sterberaten vorhanden sein müssen, um alle Daten mit in die Modellierung aufnehmen zu können. Diese Problematik zeigte sich auch in unserem Datensatz in der Überlegung, den Kohorteneinfluss mit dem durchschnittlichen BMI im Jugendalter zu ergänzen, wodurch ungefähr nur die Hälfte der Daten zur Modellierung verwendet werden hätte können. Es wurde auch überlegt, andere mögliche Risikofaktoren für Krebs im Ver-

dauungstrakt, die die Stiftung deutsche Krebshilfe (2018) benennt, wie unter anderem Alkoholkonsum, als Einfluss in die Modelle mit aufzunehmen. Für diese konnten aber ebenfalls keine geeigneteren Daten im benötigten zeitlichen Verlauf gefunden werden.

In dieser Analyse zur Modellierung und Vorhersage durch APC-Modelle stellte sich, wie in vielen Bereichen der Statistik, der Umgang mit unzureichender Datendokumentation in der Vergangenheit als Herausforderung dar. Durch die Verwendung von geschätzten Daten wird die Unsicherheit in den Modellierungen vergrößert. Die voranschreitende Digitalisierung in den letzten Jahren und die Etablierung von Big-Data-Systemen geht jedoch mit einer Verbesserung in der Breite und Präzision der Datendokumentation einher. Das kann sich in der Zukunft ebenfalls positiv auf die Genauigkeit von Prädiktionen mit Alters-Perioden-Kohorten-Modellen auswirken.

# Literaturverzeichnis

- [Berzuini und Clayton 1994] BERZUINI, C. ; CLAYTON, D.: Bayesian Analysis of Survival on Multiple Time Scales. In: *Statistics in Medicine* 13 (1994), S. 823–838
- [Berzuini u. a. 1993] BERZUINI, C. ; CLAYTON, D. ; BERNADINELLI, L.: Bayesian Inference on the Lexis Diagram. In: *Bulletin of the International Statistical Institute* 13 (1993), S. 823–838
- [Besag u. a. 1995] BESAG, J. ; GREEN, P. ; HIGDON, D. ; Mengersen, K.: Bayesian Computation and Stochastic Systems. In: *Statistical Science* 10 (1995), S. 3–66
- [Boeing u. a. 1991] BOEING, H. ; FRENTZEL-BEYME, R. ; BERGER, M. ; U.A.: Case-control Study on Stomach Cancer in Germany. In: *International Journal of Cancer* 47 (1991), S. 858–864
- [Carter und Kohn 1994] CARTER, C. K. ; KOHN, R.: On Gibbs Sampling for State-Space Models. In: *Biometrika* 81 (1994), S. 541 – 553
- [Deutsches Institut für Medizinische Dokumentation und Information DIMDI 2018] DEUTSCHES INSTITUT FÜR MEDIZINISCHE DOKUMENTATION UND INFORMATION DIMDI: ICD-10-GM Version 2019, Systematisches Verzeichnis, Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme. In: *WHO Technical Report Series* 10 (2018). – URL [www.dimdi.de](http://www.dimdi.de)
- [Ebert 2018] EBERT, Thomas: Die Zukunft des Generationenvertrags. In: *Schriftenreihe* 10293 (2018). – URL [www.bpb.de](http://www.bpb.de)
- [Frühwirth-Schnatter 1994] FRÜHWIRTH-SCHNATTER, S.: Data Augmentation and Dynamic linear Models. In: *Journal of Time Series Analysis* 15 (1994), S. 183–202
- [Holford 1983] HOLFORD, T.R.: The Estimation of Age, Period and Cohort Effects for Vital Rates. In: *Biometrics* 39 (1983), S. 311–324
- [Knorr-Held und Rainer 2001] KNORR-HELD, L. ; RAINER, E.: Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction. In: *Biostatistics* 2 (2001), S. 109–129



- [NCD Risk Factor Collaboration 2019] NCD RISK FACTOR COLLABORATION : *Body-Mass-Index: Evolution of Bodymass all over Time*. 2019. – URL [www.ncdrisc.org](http://www.ncdrisc.org)
- [R Core Team 2018] R CORE TEAM: *R: A Language and Environment for Statistical Computing*. 2018
- [Riebler und Held 2017] RIEBLER, A. ; HELD, L.: Projecting the Future Burden of Cancer: Bayesian Age–Period–Cohort Analysis with Integrated nested Laplace Approximations. In: *Biometrical Journal* 57 (2017), S. 531–549
- [Schmid und Held 2004] SCHMID, V. ; HELD, L.: Bayesian Extrapolation of Space-Time Trends in Cancer Registry Data. In: *Biometrics* 60 (2004), S. 1034–1042
- [Schmid und Held 2007] SCHMID, V. ; HELD, L.: Bayesian Age-Period-Cohort Modeling and Prediction - BAMP. In: *Journal of Statistical Software* 21 (2007)
- [Spiegelhalter u. a. 2002] SPIEGELHALTER, D.J. ; BEST, N.G. ; CARLIN, B.P. ; LINDE, A. van der: Bayesian Measures of Model Complexity and Fit. In: *Journal of the Royal Statistical Society B* 64 (2002), S. 583–639
- [Statistisches Bundesamt DeStatis 2019] STATISTISCHES BUNDESAMT DESTATIS: *Indikator 3.2 des Indikatorensatzes der GBE der Länder: Sterbefälle (Anzahl, je 100.000 Einwohner, je 100.000 Einwohner -altersstandardisiert an der Alten Europastandardbevölkerung-) nach Hauptdiagnosegruppen und Geschlecht, Region, ab 1998*. 2019. – URL [www.gbe-bund.de](http://www.gbe-bund.de)
- [Stiftung deutsche Krebshilfe 2018] STIFTUNG DEUTSCHE KREBSHILFE: *Gesunden Appetit! Vielseitig essen - gesund leben*. 2018. – URL [www.krebshilfe.de](http://www.krebshilfe.de)
- [WHO 2000] WHO: Obesity: Preventing and Managing the global Epidemic. In: *WHO Technical Report Series* 894 (2000)

## Tabellenverzeichnis

3.1	Posterior-Median der Devianz, effektive Anzahl an Parametern und DIC für alle Modelle. . . . .	15
4.1	Anteile beobachteter Fälle innerhalb der vorhergesagten Kreditibilitätsintervalle und Devianz des vorhergesagten Medians. . . . .	21

# Abbildungsverzeichnis

3.1	Todesfälle je Altersgruppe (oben) und Sterblichkeitsrate als Todesfälle je 100 000 je Altersgruppe (unten) nach Geschlecht. . . . .	10
3.2	Posterior geschätzte RW 1 Effekte für Alter, Periode und Kohorte mit Median und 90% Kredibilitätsintervall im Modell ohne Überdispersion. 12	
3.3	Posterior geschätzte RW 1 Effekte für Alter, Periode und Kohorte mit Median und 90% Kredibilitätsintervall im Modell mit Überdispersion. 13	
3.4	Altersstandardisierter durchschnittlicher BMI für Erwachsene und Kinder zwischen 5 und 19 Jahren in Deutschland. . . . .	17
4.1	Beobachtete und ab 2012 vorhergesagte Sterbefälle mit Median und 90% Kredibilitätsintervall aggregiert über alle Altersgruppen; für Männer (oben) und Frauen (unten) mit RW 1(links) und RW 2 (rechts) ohne Überdispersion. . . . .	22
5.1	Beobachtete und vorhergesagte Sterberaten als Sterbefälle je 100 000 mit Median und 90% Kredibilitätsintervall; für Männer in Altersgruppen 50 bis 69 Jahre mit RW 1(links) und RW 2 (rechts) ohne Überdispersion. . . . .	25
5.2	Beobachtete und vorhergesagte Sterberaten als Sterbefälle je 100 000 mit Median und 90% Kredibilitätsintervall; für Männer in Altersgruppen 70 bis 89 Jahre mit RW 1(links) und RW 2 (rechts) ohne Überdispersion. . . . .	26
5.3	Beobachtete und vorhergesagte Sterberaten als Sterbefälle je 100 000 mit Median und 90% Kredibilitätsintervall; für Frauen in Altersgruppen 50 bis 69 Jahre mit RW 1(links) und RW 2 (rechts) ohne Überdispersion. . . . .	27
5.4	Beobachtete und vorhergesagte Sterberaten als Sterbefälle je 100 000 mit Median und 90% Kredibilitätsintervall; für Frauen in Altersgruppen 70 bis 89 Jahre mit RW 1(links) und RW 2 (rechts) ohne Überdispersion. . . . .	28

# Eigenständigkeitserklärung

Ich versichere, dass ich die vorgelegte Bachelorarbeit eigenständig und ohne fremde Hilfe verfasst, keine anderen als die angegebenen Quellen verwendet und die den benutzten Quellen entnommenen Passagen als solche kenntlich gemacht habe. Diese Bachelorarbeit ist in keinem anderen Kurs in dieser oder einer ähnlichen Form vorgelegt worden.

Name, Vorname: Schweisthal, Jonas

München, den 21.08.2019

Unterschrift: